DOCUMENT RESUME

ED 452 233 TM 032 519

AUTHOR	Hombo, Catherine M.; Pashley, Katharine; Jenkins, Frank
TITLE	Are Grid-In Response Format Items Usable in Secondary Classrooms?
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	2001-04-00
NOTE	14p.
CONTRACT	R999G50001
PUB TYPE	Reports - Research (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Responses; Secondary Education; *Secondary School Students; *Standardized Tests; Test Format; *Test Items
IDENTIFIERS	National Assessment of Educational Progress

ABSTRACT

The use of grid-in formats, such as those requiring students to solve problems and fill in bubbles, is common on large-scale standardized assessments, but little is known about the use of this format with a more general population of students than high school students taking college entrance examinations, including those attending public schools in grades 8 and 12. Data were taken from the National Assessment of Educational Progress (NAEP) mathematics field test for 2,673 examinees in grade 8 and 2,793 in grade 12. It was expected that there would be effects from requiring the grid-in format, especially for items that were regularly multiple choice, but the effects of the grid-in format are larger than expected for short constructed response items. Students were also inconsistent in what they wrote in the blocks and what they gridded below the blocks, resulting in a correct response in one format and not in the other. The presence of the grid-in response format items apparently discouraged some students from attempting items, and this effect seems to extend through the blocks containing the grid-in items to other items. The results indicate that substantial student practice and familiarity with the format will be a necessary condition to the use of such items. (Contains 12 tables.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 452 233

Are Grid-In Response Format Items Usable In Secondary Classrooms?

Catherine M. Hombo
Katharine Pashley
Frank Jenkins

Educational Testing Service

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM032519

The work reported herein was supported under the National Assessment of Educational Progress (Grant No. R999G50001, CFDA No. 84.999G) as administered by the Office of Educational Research and Improvement, U.S. Department of Education. The opinions expressed herein are solely those of the authors and do not necessarily represent those of Educational Testing Service.

BEST COPY AVAILABLE

Introduction

Many classroom teachers are interested in expanding their repertoire of item formats when creating classroom assessment instruments. Items that can be machine-scored are especially useful for assessments with tight deadlines to report grades for students, such as end-of-semester exams. However, traditional machine-scorable item formats such as multiple-choice, matching, and true/false are not always the best item types to assess the content and skills appropriate to the exam. Especially in mathematics, the ability to require solutions to complex and demanding items in a machine-scorable response format would be very desirable.

One possibility is items presented in a grid-in format. Grid-in items require examinees to solve an item in a free-response format, then fill in bubbles representing the solution's digits and possibly a decimal point in a provided grid. Such item formats now make an appearance on large-scale, high-stakes assessments such as the SAT and the GRE. The populations of students who take such assessments are generally high-ability, relatively mature students, often with substantial advance knowledge of the test and item format. Grid-in item formats have been examined with these populations of examinees and found to perform quite well. However, little is known about the performance of a more general population of students, such as those attending public schools in grades 8 and 12.

Source of the data

The National Assessment of Educational Progress (NAEP) is a government-mandated survey of the educational achievement of American students in reading, mathematics, science, and writing. The student samples for NAEP assessments are carefully chosen to represent, as closely as possible, the population of US students in school at the grade level and subject area being assessed. Nearly one-third of the items used in NAEP assessments are publicly released and must be replaced by new items. The properties of the new items are evaluated in field tests (U.S. Department of Education, 1999).

In the 1999 NAEP field test, identical item stems with more than one response format were included in mathematics at 8th and 12th grades. Items were originally presented in one of two regular response formats, either multiple-choice (MC), short constructed response (SCR), OR Extended Constructed Response (ECR; grade 12 only). The constructed response items were structured so that the examinee response was numeric and could be gridded in the alternate format. Items were presented to examinees in the study in two ways, their regular format and grid-in. Grid-in items required examinees to both write their responses in the provided boxes at the top of the grid and fill in bubbles representing the digits and possibly a decimal point in a provided, four-block grid. The booklets were structured so that each examinee saw a mix of item response formats across items, but no single item in more than one response format.

Research Questions

It was anticipated that differential examinee performance might be observed across several levels of analysis:

1. Difference in item difficulty between the regular response format for the items (either multiple-choice or constructed response) and grid-in response formats
2. Difference in student performance between the MC-grid pairs and the SCR-grid pairs
3. Difference in omit and non-response rates when grid-in items are placed together in a group in contrast to placement scattered throughout the block
4. Differential speededness of blocks of items containing the grid-in items and those presented in multiple-choice and short constructed response formats

Method

The grid-in response format has 4 boxes in which students are expected to write either a digit or a decimal point. Beneath each box there is a column containing the possible grid-in choices for that blank. The first column has the digits 1 through 9 and a decimal point. The other three columns have the digits 0 through 9 and a decimal point.

Grid-in items were included in eight item blocks; four presented at the each of the 8th and 12th grade levels. See Table 1 for information about the structure of the blocks and item response formats in grade 8, and Table 2 for grade 12. Blocks MX7 and MX9 consist of the same items, as do blocks MX8 and MX10, but the presentation format alternates. To simplify the presentation of results, the items that make up blocks MX7/MX9 will be referred to as block 1, and those forming blocks MX8/MX10 will be referred to as block 2.

The item blocks at the 8th grade consisted of 8 items and at 12th grade consisted of 11 items, and students are allowed 15 minutes to complete a block. Some blocks included one extended constructed response item. The grid-in items were placed in the blocks in three ways, 1) alternating with the regular-format items or grouped together as the 2) first or 3) last items. There were 2673 examinees at the 8th grade, including 1411 females and 1262 males, and 2793 at the 12th grade, including 1388 females and 1405 males. Approximately 60% of students in both samples were White.

There were three types of correct grid-in answers: a single number (27 items), any one of multiple correct responses (4 items), or a continuous finite interval (2 items). To score the grid-in items, a list of "correct" responses had to be generated. Allowances were made for collapsing of blank spaces, leading zeroes, and trailing decimal characters in student responses with no penalty. These were believed to be due to unfamiliarity with the response format and not related to the trait being measured. One limitation to the use of the grid-in format is the necessity of making somewhat subjective decision about exactly which responses will be scored as correct.

For example, an item with stem "6 + 4 =" would have keyed response "10" with the following responses accepted as correct:

10_ _
_10_
_ _10
10..
10._
_010
10.0

where "_" represents a blank in which no character was written or gridded. Decisions must be made about whether or not to treat responses such as _1_0 as correct. For the purposes of this study, only the seven responses listed above were accepted as correct. There are other possible "correct" responses, such as 0010, but they did not occur in the response data.

Two parts of the student grid-in responses were scored: the gridded and the written. The gridded part was read and scored by standard scoring software. The written part appeared in the blocks at the top of the grid and was read using Intelligent Character Recognition (ICR) software.

It was decided that the ICR results were insufficiently accurate to use in the analysis. Table 3 presents some information regarding the accuracy of the ICR scanning for the items and blocks administered to the grade 8 examinees. In Table 3, "Matched Responses" indicates that the ICR and the grid-in results were the same for an examinee on an item. The accuracy of the software was enhanced by human verification and, when necessary, key entry of illegible responses and revision of incorrectly read data, before the ICR data was used in the analysis.

Results

As can be seen in Table 3, the ICR results are unacceptably inaccurate for general use. The mean correct capture rate across items is 72.35%, indicating a loss of more than a quarter of the information. While this problems can be remedied by visual checking and key entry of inaccurately-scanned entries, this defeats the purpose of using ICR and adds and expensive and time-consuming layer of effort.

Item difficulty results for grade 8 are presented in Table 4, and for grade 12 in Table 5. Item difficulty varied considerably across response formats for several items. There was sizable mismatch between the response formats within the grid-in items, between the student's written response and the gridded response, as noted above. There were differences in performance between the MC-grid pairs and the SCR-grid pairs, but the differences were smaller than expected.

One factor was not accounted for in the assessment design, the apparent increase in testing time required for completion of the grid-in items. The percentage of missing data for each item is presented in Table 6 for grade 8 and Table 7 for grade 12. There were large apparent differences in speededness between blocks with and without grid-in response format items. Non-response rates for the last item in a block containing grid-in response format items were greater than 60%, much larger than rates typically seen in NAEP assessments. Generally, grid-in format items were omitted substantially more often than multiple-choice format versions of the same items, and more often than constructed response format versions of the same items. Of course, item difficulty and item format impact omit rates, as more difficult items and items requiring an extended constructed response tend to have higher omit rates under any circumstances.

Item response format placement within block also had a noticeable effect on student performance and omit frequency. Blocks of grid-in format items seemed to produce greater omit rates throughout their block than when the grid-in and regular response formats were mixed throughout the block. Overall block omit rates were highest when the grid-in items were clustered at the beginning of the block of items.

Discussion

It is apparent from this study that seemingly small changes in items can have considerable impact on student performance. With items that were regularly multiple-choice response format items, the requirement that the examinee produce an answer rather than choose one of offered alternatives seemed to indicate that large effects could be expected. With items that were regularly short constructed response format items, students had to produce and write an answer to the item in both formats. In this case, the impact of requiring use of the grid-in response format was not expected to be as large as the results of the study indicated.

Another disturbing finding is the mismatch between the written and gridded response within the grid-in item response format. Students were inconsistent in what they wrote in the blocks and what they gridded below the blocks, resulting in a correct response in one format but not the other. This occurred more frequently than anticipated. This effect may be due to unfamiliarity with the format, and possibly would be alleviated with practice, but it sounds a warning for users of new, free-response formats: scoring of items in these formats may be strongly influenced by factors other than ability and subject-area knowledge.

The presence of grid-in response format items apparently discourages some students from attempting items, especially those in the grid-in response format. This effect seemed to extend throughout the blocks containing the grid-in items to items in other response formats. The grid-in items seemed to increase the speededness of the blocks. This may be due to the unfamiliarity of the format, or to the fact that gridding a response takes longer than simply writing it down, or some combination of these factors.

Given the increasing interest and popularity of non-multiple-choice format assessments, care must be taken that the quality of the assessment are maintained when new presentation formats are developed and used. It is possible that machine-scorable, grid-in format items have a place in classroom assessment, but substantial student practice and familiarity with the format will be a necessary condition to such use. In addition, increases in the accuracy of ICR recovery would aid in increasing use of this technology.

References

U.S. Department of Education. Office of Research and Improvement. National Center for Educational Statistics. The NAEP 1996 Technical Report, NCES 1999-452, by Allen, N. L., Carlson, J. E., & Zelenak, C. A. (1999). Washington, DC: National Center for Educational Statstics.

Table 1
Block Structure - Grade 8

Item Number	Block	Item Type / Response Format	Block	Item Type / Response Format
1	MX7	MC / Regular	MX9	SCR / Grid-in
2	MX7	SCR / Grid-in	MX9	MC / Regular
3	MX7	MC / Regular	MX9	SCR / Grid-in
4	MX7	SCR / Grid-in	MX9	MC / Regular
5	MX7	MC / Regular	MX9	SCR / Grid-in
6	MX7	MC / Regular	MX9	SCR / Grid-in
7	MX7	SCR / Grid-in	MX9	MC / Regular
8	MX7	SCR / Grid-in	MX9	MC / Regular
1	MX8	SCR / Grid-in	MX10	MC / Regular
2	MX8	SCR / Grid-in	MX10	MC / Regular
3	MX8	SCR / Grid-in	MX10	MC / Regular
4	MX8	SCR / Grid-in	MX10	MC / Regular
5	MX8	MC / Regular	MX10	SCR / Grid-in
6	MX8	MC / Regular	MX10	SCR / Grid-in
7	MX8	MC / Regular	MX10	SCR / Grid-in
8	MX8	MC / Regular	MX10	SCR / Grid-in

NOTE: Items in the same row have identical stems and differ only in response format.

Table 2

Block Structure - Grade 12

Item Number	Block	Item Type / Response Format	Block	Item Type / Response Format
1	MX7	SCR / Grid-in	MX9	MC / Regular
2	MX7	MC / Regular	MX9	SCR / Grid-in
3	MX7	SCR / Regular	MX9	SCR / Grid-in
4	MX7	SCR / Grid-in	MX9	ECR / Regular
5	MX7	MC / Regular	MX9	SCR / Grid-in
6	MX7	SCR / Grid-in	MX9	SCR / Regular
7	MX7	MC / Regular	MX9	SCR / Grid-in
8	MX7	SCR / Grid-in	MX9	SCR / Regular
9	MX7	SCR / Regular	MX9	SCR / Grid-in
10	MX7	SCR / Grid-in	MX9	MC / Regular
11	MX7	SCR / Regular	MX9	SCR / Grid-in
1	MX8	SCR / Grid-in	MX10	SCR / Regular
2	MX8	SCR / Regular	MX10	SCR / Grid-in
3	MX8	SCR / Grid-in	MX10	MC / Regular
4	MX8	SCR / Grid-in	MX10	MC / Regular
5	MX8	SCR / Grid-in	MX10	MC / Regular
6	MX8	ECR / Regular	MX10	SCR / Grid-in
7	MX8	SCR / Regular	MX10	SCR / Grid-in
8	MX8	MC /Regular	MX10	SCR / Grid-in

NOTE: Items in the same row have identical stems and differ only in response format.

Table 3

Accuracy of ICR scanning – Grade 8 Blocks / Items

Item	Matched Responses	Total Correct Responses	Percent of Correct Captures
1	365	508	71.85
2	371	518	71.62
3	373	523	71.32
4	333	457	72.87
5	403	557	72.35
6	380	533	71.29
7	409	550	74.36
8	322	456	70.61
1	452	568	79.58
2	340	495	68.69
3	382	516	74.03
4	353	504	70.04
5	371	510	72.75
6	288	423	68.09
7	305	410	74.39
8	266	361	73.68

Table 4

Proportion Correct Results – Grade 8, Blocks 1 and 2

Item Number	Standard Format	Grid-in response	ICR response	Overall
1	0.401	0.268	0.234	0.308
2	0.778	0.752	0.706	0.746
3	0.416	0.318	0.279	0.342
4	0.355	0.167	0.138	0.233
5	0.623	0.696	0.631	0.646
6	0.618	0.592	0.584	0.599
7	0.797	0.730	0.707	0.746
8	0.225	0.176	0.158	0.189
1	0.765	0.764	0.743	0.757
2	0.232	0.041	0.041	0.111
3	0.494	0.244	0.221	0.326
4	0.591	0.553	0.523	0.557
5	0.437	0.132	0.116	0.232
6	0.218	0.150	0.137	0.172
7	0.575	0.388	0.374	0.450
8	0.084	0.018	0.017	0.041

Table 5

Proportion Correct Results – Grade 12, Blocks 1 and 2

Item Number	Standard Format	Grid-in response	ICR response	Overall
1	0.640	0.507	0.467	0.542
2	0.635	0.565	0.554	0.586
3	0.327	0.351	0.316	0.331
4	0.250	0.285	0.268	0.268
5	0.150	0.114	0.115	0.128
6	0.585	0.617	0.577	0.592
7	0.477	0.271	0.259	0.340
8	0.252	0.422	0.385	0.346
9	0.132	0.160	0.142	0.144
10	0.190	0.195	0.166	0.184
11	0.060	0.053	0.043	0.052
1	0.266	0.302	0.282	0.283
2	0.453	0.552	0.540	0.514
3	0.370	0.187	0.181	0.251
4	0.340	0.324	0.287	0.318
5	0.398	0.480	0.454	0.440
6	0.097	0.147	0.135	0.126
7	0.111	0.108	0.101	0.107
8	0.206	0.042	0.044	0.100

Table 6

Grade 8 Percent Missing Data Rates – Blocks 1 and 2

Item Number	Standard Format	Grid-in response
1	1.92	11.18
2	1.23	11.54
3	1.18	9.34
4	2.60	22.63
5	1.63	3.83
6	2.81	7.96
7	1.99	22.19
8	8.73	15.24
1	0.77	4.44
2	8.42	16.57
3	4.75	12.28
4	4.44	14.35
5	15.68	26.34
6	20.71	28.33
7	30.03	35.99
8	36.39	48.39

Table 7

Grade 12 Percent Missing Data Rates – Blocks 1 and 2

Item Number	Standard Format	Grid-in response
1	6.11	13.30
2	2.15	7.24
3	9.87	11.93
4	34.09	26.32
5	13.16	29.55
6	14.77	18.88
7	19.31	25.00
8	36.79	44.21
9	38.34	45.17
10	43.18	55.22
11	57.37	60.80
1	10.37	11.59
2	7.44	8.81
3	11.65	18.45
4	8.10	20.31
5	5.68	25.89
6	16.02	28.69
7	22.17	20.03
8	45.06	25.57